Invariant theory and geodesically convex optimization

Ankit Garg
Microsoft Research New England

Based on joint works with
Zeyuan Allen-Zhu, Peter Bürgisser,
Leonid Gurvits, Yuanzhi Li, Rafael Oliveira,
Michael Walter and Avi Wigderson

Overview

- Introduction to algorithmic problems in *invariant theory*.
- *Non-convex* optimization problems but *geodesically convex*.
- *Connections* to several areas of computer science, mathematics and physics.

Geometric complexity theory – asymptotic vanishing of Kronecker coefficients.

Quantum information theory– one-body quantum marginal problem.

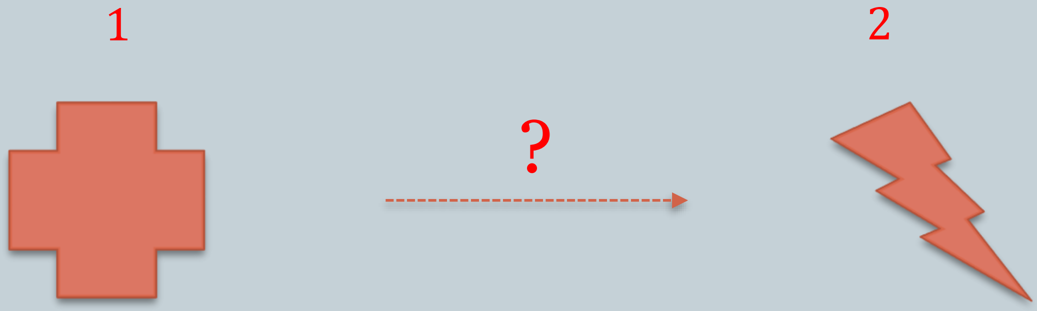
Functional analysis – Brascamp-Lieb inequalities.

Optimization– Geodesic convexity. Captures general linear programming.

Complexity theory and derandomization – Special cases of polynomial identity testing.

Motivating puzzle

- Given two polygons in the plane, cut 1 into a finite number of pieces and rearrange to get 2.



- Possible iff areas equal.
- Area is an *invariant* and the *only* invariant.

Invariant theory

- Started with [Cayley 1846]. Developed later by Hilbert, Mumford and others.
- Linear action of a group G on a vector space V.

Example 1

- $G = S_n$ acts on $V = C^n$ by permuting coordinates.

$$\sigma \cdot (x_1, \dots, x_n) \to (x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

- Symmetric polynomials are invariant under the group action.
- Generated by the n elementary symmetric polynomials.

Example 2

- $G = SL_n(C) \times SL_n(C)$ acts on $V = M_n(C)$ by left-right multiplication.

$$(A, B) \cdot X = AXB.$$

- $\mathrm{Det}(X)$ is invariant.
- Every polynomial invariant of the form $q(X) = p(\mathrm{Det}(X))$.

Invariant theory

- Underlying field complex numbers C.
- Groups G: *finite*, $GL_n(C)$, $SL_n(C)$, *direct products* of these etc.

G acts (linearly) on V

- $\mathrm{id}\cdot v = v$ for all $v \in V$.
- $(g_1 g_2) \cdot v = g_1 \cdot (g_2 \cdot v)$ for all $g_1 . g_2 \in G, v \in V$.
- $g \cdot (cv) = cg \cdot v$ for all $c \in C, v \in V$.
- $g \cdot (v_1 + v_2) = g \cdot v_1 + g \cdot v_2$ for all $g \in G, v_1, v_2 \in V$.

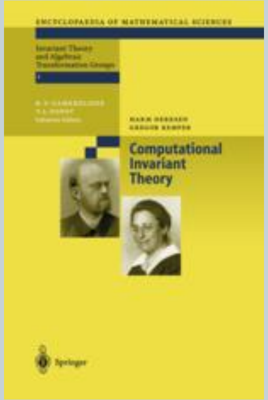
Objects of study

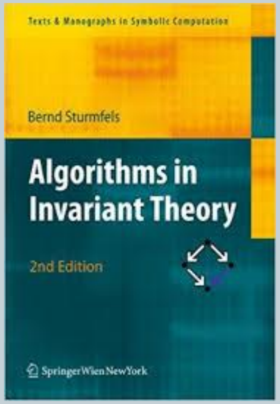
- Invariant polynomials: Polynomial functions on V invariant under action of G. p s.t. $p(g \cdot v) = p(v)$ for all $g \in G, v \in V$.
- Orbits: Orbit of vector $v = \{g \cdot v : g \in G\}$.
- Orbit-closures: Orbits may not be closed. Take their closures.

Orbits and orbit-closures

- Capture several interesting problems in theoretical computer science.
- *Graph isomorphism*: Whether orbits of two graphs the same. Group action: permuting the vertices.
- *Arithmetic circuits*: The VP vs VNP question. Whether permanent lies in the orbit-closure of the determinant. Group action: Action of $GL_{n^2}(C)$ on polynomials induced by action on variables.
- *Tensor rank*: Whether a tensor lies in the orbit-closure of the diagonal unit tensor. Group action: Natural action of $GL_n(C) \times GL_n(C) \times GL_n(C)$.

Computational invariant theory





- Highly *algorithmic field*.
- Algorithms sought and well developed.
- Algorithms still remain *exponential time* or even doubly exponential time in some cases.
- Bottleneck: *Gröbner basis* computation or more generally *algebraic methods*.

Null cone

- Fix an action of G on V.
- Null cone: Vectors v s.t. 0 lies in the orbit-closure of v.
- $\{v: 0 \in \overline{G \cdot v}\}$.
- Sequence of group elements $g_1, \ldots, g_k, \ldots$ s.t. $\lim_{k \to \infty} g_k \cdot v = 0$.
- Problem: Given $v \in V$, decide if it is in the null cone.
- Will capture many interesting questions.

- [Hilbert 1893; Mumford 1965]: v in null cone iff $p(v) = 0$ for all homogeneous invariant polynomials p.
- One direction clear (polynomials are continuous).
- Other direction uses *Nullstellansatz* and some algebraic geometry.

Example 1

- $G = SL_n(C) \times SL_n(C)$ acts on $V = M_n(C)$ by left-right multiplication.

$$(A, B) \cdot X = AXB.$$

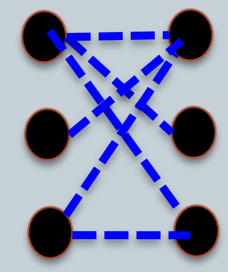
- $\text{Det}(X)$ is invariant.
- Every polynomial invariant of the form $q(X) = p(\text{Det}(X))$.
- Null cone: Singular matrices.

Example 2

- ST_n: group of $n \times n$ diagonal matrices with determinant 1.
- $G = ST_n \times ST_n$ acts on $V = M_n(C)$ by left-right multiplication.

$$(A, B) \cdot X = AXB.$$

- $\text{Det}(X)$ is invariant.
- Are there other invariants?
- $X_{1,1} \cdot X_{2,2} \cdots X_{n,n}$ is invariant.
- So is $X_{1,\sigma(1)} \cdot X_{2,\sigma(2)} \cdots X_{n,\sigma(n)}$ for any permutation σ. And these are all.
- Null cone: perfect matching.
- A_H is in null cone iff H has no perfect matching.



H

1	1	1
1	0	0
1	0	1

A_H

Example 3: Linear programming

- T_n: group of $n \times n$ diagonal matrices.
- $G = T_n$ acts on $V = C^m$. $\boldsymbol{t} \in T_n$, $\boldsymbol{t} = \mathrm{diag}(t_1, \dots, t_n)$.
- Fix m vectors $w^{(1)}, \dots, w^{(m)} \in C^n$.

$$\boldsymbol{t} \cdot e_j = \prod_{i=1}^{n} t_i^{w_i^{(j)}}\ e_j.$$

- $v \in V$. Define $\mathrm{supp}(v) = \{j \in [m] : v_j \neq 0\}$.
- Theorem: *v not in* null cone iff *convex hull* of $\left\{w^{(j)} : j \in supp(v)\right\}$ contains 0.
- Captures linear programming.
- So the null cone (membership) problem is a *non-commutative* analogue of *linear programming*.

Symbolic matrices

- Symbolic matrices: $L = \sum_{i=1}^{m} x_i A_i$.
- $A_1, \ldots, A_m$ are $n \times n$ matrices over C.
- $x_1, \ldots, x_m$ are formal variables (commuting or non-commuting).
- L has entries linear polynomials in $x_1, \ldots, x_m$.

- Question: Is L singular?
- Dual life depending on whether the variables commute or not.

Commuting variables

- $L = \sum_{i=1}^{m} x_i A_i$.
- SING: Is L singular over $C(x_1, \dots, x_m)$? [Edmonds' problem 67].
- $C(x_1, \dots, x_m)$ - field of rational functions.
- Proposition: L SING if $\sum_i \alpha_i A_i$ singular for all $\alpha_1, \dots, \alpha_m \in C$.
  (all matrices in the subspace spanned by $A_1, \dots, A_m$ singular).
- Efficient randomized algorithm: plug in random values for $x_1, \dots, x_m$ [Lovász 79].
- Efficient deterministic algorithm – *major open problem*.
- Captures polynomial identity testing (*PIT*) [Valiant 79].
- Implies circuit lower bounds [Kabanets, Impagliazzo 04].

Non-commuting variables

- $L = \sum_{i=1}^{m} x_i A_i$.
- NC-SING: Is L singular for non-commuting $x_1, \ldots, x_m$?
- Various ways to define. Extensive work by Amitsur, Cohn and others.
- L NC-SING if for all d, $\sum_i B_i \otimes A_i$ singular for all $B_1, \ldots, B_m \in M_d(C)$.
- Plug in matrices for x_i's. Take *"matrix linear combinations"* instead of linear combinations.

Example

- SING and NC-SING can differ.
- $L = \begin{pmatrix} 0 & x & y \\ -x & 0 & z \\ -y & -z & 0 \end{pmatrix}$.
- Skew symmetric. When x, y, z commute, L is singular.
- Homework: When x, y, z non-commuting, then L is not singular.

Example 4: Noncommutative singularity

- $G = SL_n(C) \times SL_n(C)$ acts on $V = M_n(C)^{\oplus m}$ by *simultaneous left-right* multiplication.

$$(B, C) \cdot (A_1, \ldots, A_m) = (BA_1C, \ldots, BA_mC).$$

- Theorem [Derksen, Weyman 00, …]: $(A_1, \ldots, A_m)$ in null cone iff $L = \sum_i x_i A_i$ NC-SING.
- Theorem [G, Gurvits, Oliveira, Wigderson 16]: Deterministic polynomial time algorithm for NC-SING.
- Brascamp-Lieb inequalities: Generalize *Cauchy-Schwarz*, *Hölder's*, *Loomis-Whitney*, *Shearer's* inequalities.
- Feasibility can be phrased as a null cone problem.
- Theorem [G, Gurvits, Oliveira, Wigderson 16b]: Polynomial time algorithm to test feasibility and compute optimal constants.

Null cone and optimization

- Recap: Null cone membership unifies many problems in computer science and mathematics.
- G acting on V.
- $v \in V$ in null cone if there is sequence of group elements $g_1, \ldots, g_k, \ldots$ s.t. $\lim_{k \to \infty} g_k \cdot v = 0$.
- Equivalently, $NC(v) = \inf_{g \in G} \|g \cdot v\|_2^2 = 0$.
- $f_v(g) = \|g \cdot v\|_2^2$. *Non-convex* but has a lot of structure.

Example 1

- ST_n: group of $n \times n$ diagonal matrices with determinant 1.
- $G = ST_n \times ST_n$ acts on $V = M_n(C)$ by left-right multiplication.

$$(B, C) \cdot A = BAC.$$

- $NC(A)$: $\inf_{\prod_i b_i = \prod_j c_j = 1} \sum_{i,j} |A_{i,j}|^2 |b_i|^2 |c_j|^2$.
- $\inf_{x,y>0, \prod_i x_i = \prod_j y_j = 1} \sum_{i,j} |A_{i,j}|^2 x_i y_j$.
- *Non-convex* but a simple transformation makes it convex.
- $x_i = \exp(\lambda_i)$ and $y_j = \exp(\mu_j)$ makes it convex.

Example 2

- $G = SL_n(C) \times SL_n(C)$ acts on $V = M_n(C)^{\oplus m}$ by *simultaneous left-right* multiplication.

$$(B, C) \cdot (A_1, \dots, A_m) = (BA_1C, \dots, BA_mC).$$

- $NC(A_1, \dots, A_m)$: $\inf_{X,Y \succ 0, \mathrm{Det}(X)=\mathrm{Det}(Y)=1} \ \mathrm{tr}\left[\sum_i A_i X A_i^\dagger Y\right]$.
- *Non-convex*. No change of variables that makes it convex.
- $X = \exp(P), Y = \exp(Q)$ doesn't work.
- *Non-commutativity* of matrix multiplication. $\exp(M + N) \neq \exp(M)\exp(N)$.

Alternating minimization for NC-SING

- Recall: $A_1, \dots, A_m$ NC-SING if

$$\mathrm{Det}\left(\sum_{i=1}^{m} B_i \otimes A_i\right) = 0,$$

for all d, for all B_i ($d \times d$ matrices).

- Also iff $NC(A_1, \dots, A_m) > 0$.
- [KKT condition]: $NC(A_1, \dots, A_m) > 0$ iff can be *transformed* to satisfy the following two conditions:

$$\sum_i A_i A_i^T \approx I \text{ and } \sum_i A_i^T A_i \approx I.$$

- Allowed: *Simultaneous left-right* multiplication by matrices.
- $A_1, \dots, A_m \to BA_1C, \dots, BA_mC$.

Alternating minimization for NC-SING

- Goal: Transform $A_1, \ldots, A_m$ to satisfy
  $$\sum_i A_i A_i^T \approx I \text{ and } \sum_i A_i^T A_i \approx I.$$
- Left normalize: $A_1, \ldots, A_m \to \left(\sum_i A_i A_i^T\right)^{-1/2} A_1, \ldots, \left(\sum_i A_i A_i^T\right)^{-1/2} A_m$.
- Ensures $\sum_i A_i A_i^T = I$.
- Right normalize: $A_1, \ldots, A_m \to A_1\left(\sum_i A_i^T A_i\right)^{-1/2}, \ldots, A_m\left(\sum_i A_i^T A_i\right)^{-1/2}$.
- Ensures $\sum_i A_i^T A_i = I$.

Algorithm [G, Gurvits, Oliveira, Wigderson 16]

- Repeat for n^2 steps:
1. Left normalize;
2. Right normalize;
- Test if $\sum_i A_i A_i^T \approx_{1/n} I$ and $\sum_i A_i^T A_i \approx_{1/n} I$.
  Yes: not NC-SING.
  No: NC-SING.

Analysis

Wlog assume $A_1, \dots, A_m$ integer matrices.

- If $A_1, \dots, A_m$ NC-SING, then know won't converge. No way to transform and satisfy both $\sum_i A_i A_i^T \approx I, \sum_i A_i^T A_i \approx I$ $(*)$.
- Goal: If $A_1, \dots, A_m$ not NC-SING, then converge in n^2 iterations.
- Need a potential function.
- Know: There exists d and $d \times d$ $B_1, \dots, B_m$ s.t. $\mathrm{Det}(\sum_{i=1}^m B_i \otimes A_i) \neq 0$.
- Potential function: $|\mathrm{Det}(\sum_{i=1}^m B_i \otimes A_i)|$ for a *nice* choice of $B_1, \dots, B_m$.

Analysis

- [Lower bound]: Initially $|\mathrm{Det}(\sum_{i=1}^m B_i \otimes A_i)| \geq 1$. Need $B_1, \dots, B_m$ to be integer matrices.
- [Progress per step]: If $1/n$-far from satisfying $(*)$, normalization increases $|\mathrm{Det}(\sum_{i=1}^m B_i \otimes A_i)|$ by a factor of $\exp(d/12n)$. Consequence of a robust *AM-GM* inequality. Holds for all $B_1, \dots, B_m$.
- [Upper bound]: If $A_1, \dots, A_m$ left or right normalized, $|\mathrm{Det}(\sum_{i=1}^m B_i \otimes A_i)| \leq \exp(dn \log(n))$. Need $B_1, \dots, B_m$ to have *"small"* entries.

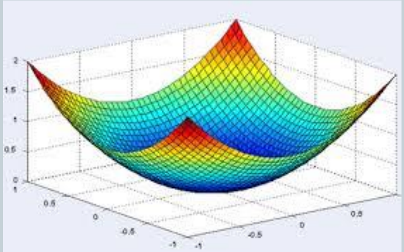
- Niceness conditions can be satisfied by using Alon's *combinatorial nullstellansatz*. Schwarz-Zippel lemma not enough.

Capacity

- Theorem [G, Gurvits, Oliveira, Wigderson 16]: Can optimize $NC(A_1 \dots, A_m)$ upto additive error ϵ in time $\text{poly}(n, \mathbf{1/\epsilon})$.
- Theorem [Allen-Zhu, G, Li, Oliveira, Wigderson 17]: Can optimize $NC(A_1 \dots, A_m)$ upto error ϵ in time $\text{poly}(n, \mathbf{\log}(\mathbf{1}/\epsilon))$.
- Crucially relies on *geodesic convexity*.
- Leads to a deterministic polynomial time algorithm for a generalization of NC-SING.

Geodesically convex optimization

- Convex optimization widely applicable.
- Works with *Euclidean geometry*.



- *Geodesic convexity*: Convexity in a different geometry.
- Lot of work on structural aspects [Udriste 94].
- Less known about algorithmic aspects.

  [Absil, Mahony, Sepulchre 09; Wiesel 12; Zhang, Sra 16].

Geodesic convex optimization

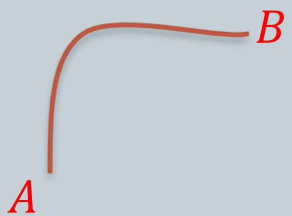
- Space of psd matrices.
- Euclidean geometry: shortest path between A and B given by straight line $\gamma(t) = tA + (1-t)B$.

- Euclidean convexity: f convex if for all shortest paths γ,

$$f(\gamma(t)) \leq t f(\gamma(0)) + (1-t) f(\gamma(1)).$$

- Extends to other geometries. Just the shortest paths (*geodesics*) change.
- Can define a geometry where the shortest path

$$\gamma(t) = B^{1/2}\left(B^{-1/2} A B^{-1/2}\right)^{t} B^{1/2}.$$



- If A and B commute, then $\gamma(t) = \exp(t \log(A) + (1-t)\log(B))$.
- Our problems are convex in this geometry!

Conclusion

- *Null cone* captures many interesting problems.
- Intriguing interplay of *analysis* and *algebra*.
- Algorithms analytic. Analysis crucially relies on algebraic character of the problem.

Open problems

- Main open problem: Design analogues of *ellipsoid/interior point* methods in the *geodesic* world (polylog$(1/\epsilon)$ dependence on error ϵ).
- Would problems in several different areas at once! In particular, the *null cone* problem.
- Concrete challenge: Null cone for the natural action of $SL_n(C) \times SL_n(C) \times SL_n(C)$ on $n \times n \times n$ tensors.
- Optimization problem: Given $v \in C^{n^3}$, test if

$$\inf_{X,Y,Z \succ 0, \mathrm{Det}(X)=\mathrm{Det}(Y)=\mathrm{Det}(Z)=1} \|(X \otimes Y \otimes Z)v\|_2^2 = 0.$$

- [Bürgisser, G, Oliveira, Walter, Wigderson 17] analyzed alternating minimization. Not enough. poly$(1/\epsilon)$ vs polylog$(1/\epsilon)$.
- Explore applications of geodesic convexity in other areas such as machine learning.

Thank You